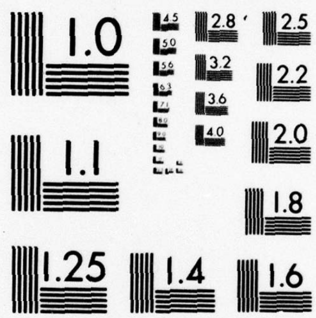
AD-A067 755 TEXAS UNIV AT AUSTIN CENTER FOR CYBERNETIC STUDIES F/G 12/1
A DUALITY THEORY FOR A CLASS OF PROBLEMS WITH ESSENTIALLY UNCON--ETC(U)
JAN 79 A BEN-TAL, Y BARZILAI, A CHARNES N00014-75-C-0616

UNCLASSIFIED CCS-330 NL

1 OF 1

AD
A067755

END
DATE
FILMED
6-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A067755

DDC FILE COPY

LEVEL II

12

CENTER FOR CYBERNETIC STUDIES

The University of Texas
Austin, Texas 78712

(12)

Research Report CCS 330

A DUALITY THEORY FOR A CLASS OF PROBLEMS WITH ESSENTIALLY UNCONSTRAINED DUALS

by

A. Ben-Tal*
Y. Barzilai*
A. Charnes**

January 1979

*Technion--Israel Institute of Technology

**The University of Texas at Austin

This research was partly supported by Project NR047-021, ONR Contracts N00014-75-C-0616 and N00014-75-C-0569 with the Center for Cybernetic Studies, The University of Texas. Reproduction in whole or in part is permitted for any purpose of the United States Government.

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
Business-Economics Building, 203E
The University of Texas at Austin
Austin, Texas 78712
(512) 471-1821

ABSTRACT

The paper introduces a class of linearly constrained convex programs whose duals are unconstrained in the sense that their solution must be in the *interior* of the feasible region. A *complete* duality theory is developed for these problems. Several examples are discussed.

Key Words: Convex Programming, Duality

A DUALITY THEORY FOR A CLASS OF PROBLEMS
WITH ESSENTIALLY UNCONSTRAINED DUALS

by

A. Ben-Tal,* Y. Barzilai,* A. Charnes**

1. Introduction

In the study of a problem arising in Statistical Information Theory [3], [1], whose mathematical model is a linearly constrained convex program:

(P) $\inf\{f(x): Ax = b\}$,

it was revealed that the dual problem is *unconstrained*. Moreover, duality states were fully characterized. Similar results can be obtained for problems arising in Transportation Engineering (estimation of Interzonal Transfers) [4], Accounting [2], and Civil Engineering (see Example 1 of this paper).

For convex programs without asympotic ideas as in [12], only partial statements can be made regarding duality states [9], [11]. Therein (i) only *sufficient* conditions are given which guarantee attainment of the primal-infimum (or the dual supremum), (ii) primal feasibility implies dual-boundedness, but is not necessarily implied by it. For *linear programs*, attainment of the primal infimum is equivalent to the *orthogonality*

* Technion - Israel Institute of Technology, Haifa, Israel
Computer Science Department.

** Center for Cybernetic Studies, University of Texas, Austin, Texas 78712, USA.

condition:

(1.1) "the price vector is orthogonal to the null space of the coefficient matrix",

and, with this, primal-feasibility is *equivalent* to dual boundedness.

In this paper we study convex programs of type (P), whose objective function belongs to a sub-class of strictly convex functions, satisfying certain smoothness properties.

For these programs, an orthogonality condition, similar to (1.1), is necessary and sufficient for the attainment of the primal infimum. Moreover, under this condition, a Slater-type constraint qualification turns out to be necessary and sufficient for attainment of the dual supremum. But the most significant consequence is that the dual problem is "unconstrained", in the sense that the optimal solution must be in the interior of the feasible set, and thus can be computed by, say, minimizing the norm of the objective function gradient.

2. Essentially Unconstrained Problems

Let f be a continuously differentiable function: $R^n \to R$, and consider the *unconstrained problem:*

(U) $\inf f(x)$.

The *solution set* of problem (U) is the (possibly empty) set

$$\Gamma_U = \{x: f(x) \leq f(y), \quad \forall y \in R^n\}.$$

For a convex problem, Γ_U is characterized simply by the critical points of f, i.e.

(2.1) $\Gamma_U = \{x: \nabla f(x) = 0\}$.

If f is defined on a subset S of R^n, the problem

$$(E) \quad \begin{cases} \inf \quad f(x) \\ \\ \text{subject to} \quad x \in S \end{cases}$$

is a *constrained* problem, and (1) generally fails to characterize its solution set.

We will term problem (E) *essentially unconstrained*, if

(i) int S is non-empty,

(ii) f is continuously differentiable on int S,

(iii) the solution set of (E) is given by

$$(2.2) \qquad \Gamma_E = \{x \in \text{int } S : \nabla f(x) = 0\},$$

where int S denotes the interior of the set S. We note that from the computational viewpoint, an essentially unconstrained problem is unconstrained. Indeed, for an unconstrained nonlinear problem, the solution set is typically the set of critical points.

A simple example of an essentially unconstrained problem is:

$$(2.3) \quad \begin{cases} \min - \sqrt{1 - x^2} \\ \\ \text{s.t.} \quad -1 \leq x \leq 1 \end{cases} .$$

The Dual of a Linearly Constrained Convex Problem

Since we are interested in linearly constrained convex programs whose duals are essentially unconstrained, we begin by introducing the *Fenchel-Rockafellar duality scheme* [9].

We recall that the *convex conjugate* f^* of $f: S \to R$, is defined by:

$$f^*(x) = \sup_{z \in S}\{<z,x> - f(z)\}$$

(where $<z,x> = \sum_{i=1}^{n} z_i x_i$ is the inner product in R^n), and the *concave conjugate* g_* of $g: S \to R$, is defined by:

$$g_*(x) = \inf_{z \in S}\{<z,x> - g(z)\} .$$

The Fenchel-Rockafellar dual problem of

$$\text{(2.4)} \qquad \inf_{x \in S}\{f(x) - g(Ax)\}$$

is

$$\text{(2.5)} \qquad \sup_{y \in R^m}\{g_*(y) - f^*(A^T y)\},$$

where f is a convex function with domain in R^n, g a concave function with domain in R^m, and A is an $m \times n$ matrix.

In terms of the indicator function

$$\delta(x|C) = \begin{cases} 0 & x \in C \\ +\infty & x \notin C \end{cases} .$$

Problem (P) is equivalent to (2.4) with

$$g(y) = -\delta(y|\{b\})$$

whose conjugate is

$$\text{(2.6)} \qquad g_*(y) = \inf_u\{<u,y> - g(u)\} = <b,y>,$$

so that (2.5) can be written as:

$$\sup_{y \in R^m} h(y) \tag{2.7}$$

with

$$h(y) = \langle b,y\rangle - f^*(A^T y) \ . \tag{2.8}$$

Note that the *Lagrangian dual* of problem (P) (see e.g., [5]) is:

$$\sup_{y \in R^m} h(y)$$

with

$$h(y) = \inf_x \{f(x) + \langle y, b-Ax\rangle\} \ . \tag{2.9}$$

However, (2.9) is equivalent to (2.8) since

$$\inf_x \{f(x) + \langle y,b - Ax\rangle\} = \langle y,b\rangle - \sup_x \{\langle A^T y,x\rangle - f(x)\} =$$

$$= \langle y,b\rangle - f^*(A^T y) \ .$$

3. Examples of Linearly Constrained Convex Problems with Essentially Unconstrained Duals

Example 1: Optimal design of a pipe network.

Consider the problem :

$$\left\{ \begin{array}{lll} \min & \sum_{j=1}^{n} L_j d_j & \\ \text{s.t.} & \sum_{j=1}^{n} \alpha_{ij} m_j d_j^{-\theta} = P_o^2 - P_i^2 & i = 1,\ldots,p \\ & d_j > 0 & j = 1,\ldots,n \ . \end{array} \right. \tag{3.1}$$

This problem arises in conjunction with the computation of the

diameters d_j, $j = 1,\ldots,n$ of n pipes in a tree-shaped fluid transportation network, so as to minimize the cost of the network, while maintaining pre-assigned minimal pressures at the end-points (see e.g. [8] and [10]).

Here L_j denotes the length of the j-th pipe, and the objective function reflects the cost of the network. The matrix (α_{ij}) describes the structure of the network as follows:

$$\alpha_{ij} = \begin{cases} 1 & \text{if pipe } j \text{ belongs to the path from the source to end-point } i. \\ 0 & \text{otherwise.} \end{cases}$$

For the j-th pipe, the pressures P_A, P_B at its end-points A, B satisfy:

$$P_A^2 - P_B^2 = \frac{m_j}{d_j^{\theta}} \tag{3.2}$$

where m_j is a constant which is independent of the diameter d_j, and θ is a dimensionless constant.

Denoting the pressure at the source and at end-point i by P_o, P_i respectively, the constraint

$$\sum_{j=1}^{n} \alpha_{ij} m_j d_j^{-\theta} = P_o^2 - P_i^2$$

relates the difference between the pressures at these two points, to the sum of pressure losses (given by (3.2)), on the path connecting them.

Problem (3.1) will be written in the form:

$$(3.3)\quad \begin{cases} \min \sum_{j=1}^{n} L_j x_j^{-\beta} \\ \text{s.t.} \quad Ax = b \\ \qquad x > 0 \end{cases}$$

where A is the $p{\times}n$ matrix whose entries are given by:

$$a_{ij} = \alpha_{ij} m_j \ ,$$

and

$$\beta = \frac{1}{\theta}$$

$$x_j = d_j^{-\theta}$$

$$b_i = P_o^2 - P_i^2 \ .$$

We note that A is a non-negative matrix, with at least one positive entry in each column (otherwise there is a pipe that is not a member of any path in the network).

We now compute the dual of a more general problem:

$$(3.4)\quad \begin{cases} \inf \sum_{i=1}^{n} c_i x_i^{-\beta_i} & c_i, \beta_i > 0 \\ \text{s.t.} \quad Ax = b \\ \qquad x \geqslant 0 \end{cases}$$

Here A is not assumed to be non-negative. However, we require that at least one row of A is positive (as is the case when one of the constraints is of the form $\sum_{i=1}^{n} x_i = 1$). Under this assumption, the dual of (3.4) is essentially unconstrained.

A direct computation shows that the conjugate of

$$q_i(x) = \begin{cases} c_i x^{-\beta_i} & x \geqslant 0 \\ +\infty & x < 0 \end{cases}$$

is

$$q_i^*(x^*) = \begin{cases} -(-x^*)^{\frac{\beta_i}{\beta_i+1}} \cdot (c_i\beta_i)^{\frac{1}{\beta_i+1}} \cdot (1 + \frac{1}{\beta_i}) & x^* \leq 0 \\ +\infty & x^* > 0 \end{cases}$$

and since $f(x) = \sum_{i=1}^{n} q_i(x_i)$:

$$f^*(A^T y) = \sum_{i=1}^{n} q_i^*(x_i^*) \quad \text{with} \quad x_i^* = \sum_{j=1}^{n} a_{ji} y_j \ .$$

Thus, the dual of (3.4) is

$$\text{(3.5)} \quad \sup \left\{ \sum_{j=1}^{m} y_j b_j + \sum_{i=1}^{n} (c_i\beta_i)^{\frac{1}{\beta_i+1}} \cdot (1 + \frac{1}{\beta_i}) \left(-\sum_{j=1}^{m} a_{ji} y_j \right)^{\frac{\beta_i}{\beta_i+1}} \right\}$$

$$\text{s.t.} \quad \sum_{j=1}^{m} a_{ji} y_j \leq 0 \qquad i = 1,\ldots,n \ .$$

Despite its formal representation as a constrained problem we shall subsequently show that (3.5) is essentially unconstrained.

<u>Example 2</u>: Let $c(y)$ be a given density function of a random variable Y. The fundamental approach of Information Theory (see [7]) to find a density which most closely resembles $c(y)$, is to find the density $x^*(t)$ (the so-called *conjugate density*, (see [6]) which minimizes the *divergence:* $\int x(y) \log \frac{x(y)}{c(y)} dy$.

Usually, $x^*(t)$ is required to satisfy constraints such as:

$$\int T(y) x^*(y) dy = \theta$$

where $T(y)$ is some statistic, and θ is, in most cases, a multi-dimensional parameter of the population.

When Y is a discrete finite random variable, we are led then to the following problem:

$$
\text{(I)} \quad \inf \sum_{j=1}^{n} x_j \log \frac{x_j}{c_j}
$$

$$
\begin{aligned}
\text{s.t.} \quad & \sum_{j=1}^{n} T_{ij} x_j = \theta_i & \quad i = 1,\dots,m \\
& \sum_{j=1}^{n} x_j = 1 & \\
& x_j \geq 0 & \quad j = 1,\dots,n
\end{aligned}
$$

or more generally

$$
(3.6) \quad \left\{
\begin{aligned}
& \inf \sum_{j=1}^{n} x_j \log \frac{x_j}{c_j} \\
& \text{s.t.} \quad Ax = b \\
& \qquad\quad x \geq 0
\end{aligned}
\right.
$$

where c_j, $j = 1,\dots,n$ are given positive scalars, A is an $m \times n$ real matrix, and b is an m-vector. Note that in the special case (Problem (I)), A has at least one positive row (corresponding to the constraint $\Sigma x_j = 1$).

The objective function $f(x) = \sum_{j=1}^{n} x_j \log \frac{x_j}{c_j}$ is separable, hence

$$
f^*(x) = \sum_{j=1}^{n} q_j^*(x) \quad \text{with} \quad q_j(x) = x_j \log \frac{x_j}{c_j} .
$$

A simple calculation shows that

$$
q_j^*(x) = \sup_{z \geq 0} \left\{ xz - z \log \frac{z}{c_j} \right\} = c_j e^{x-1} ,
$$

so that the dual of (3.6) is

$$(3.7) \qquad \sup_y \left\{ \langle y,b \rangle - \sum_{i=1}^{n} c_i \, e^{\sum_{j=1}^{m} a_{ji} y_j - 1} \right\},$$

which is an unconstrained problem, independently of the data A, b and c.

Example 3. The primal problem is

$$(3.8) \qquad \begin{cases} \inf \sum_{i=1}^{n} q_i(x_i) \\ \\ \text{s.t.} \quad Ax = b \\ \qquad\quad x \geqslant 0 \end{cases}$$

where $q_i(x_i)$ is given for $i = 1,\dots,n$ by

$$q_i(x_i) = \begin{cases} \frac{1}{2}\alpha_i x_i^2 - \beta_i \log x_i & \alpha_i,\beta_i \quad x_i > 0 \\ \\ +\infty & x_i \leqslant 0 \end{cases} .$$

The conjugate of $q_i(x_i)$ is given by

$$q_i^*(x_i^*) = \sup_{z>0} \left\{ x_i^* z - \left(\tfrac{1}{2}\alpha_i z^2 - \beta_i \log z \right) \right\} .$$

The supremum therein is attained at $\bar{z}_i$ satisfying

$$\alpha_i \bar{z}_i - \frac{\beta_i}{\bar{z}_i} = x_i^* .$$

Hence

$$\bar{z}_i = \frac{x_i^* + \sqrt{(x_i^*)^2 + 4\alpha_i\beta_i}}{2\alpha_i} ,$$

$$q_i^*(x^*) = x_i^* \frac{x_i^* + \sqrt{(x_i^*)^2 + 4\alpha_i\beta_i}}{2\alpha_i} - \frac{(x_i^* + \sqrt{(x_i^*)^2 + 4\alpha_i\beta_i})}{8\alpha_i} +$$

$$+ \beta_i \log \frac{x_i^* + \sqrt{(x_i^*)^2 + 4\alpha_i\beta_i}}{2\alpha_i} \quad ,$$

and finally, denoting by a^i the i-th column of A, the dual of (3.8) is given by

$$\sup_y \left\{ \langle y,b\rangle - \sum_{i=1}^n \left\{ \frac{1}{2\alpha_i} \left[\langle a^i,y\rangle(\langle a^i,y\rangle + \sqrt{\langle a^i,y\rangle^2 + 4\alpha_i\beta_i}) - \right.\right.\right.$$

$$\left. - \frac{1}{4}\langle a^i,y\rangle - \frac{1}{4}\sqrt{\langle a^i,y\rangle^2 + 4\alpha_i\beta_i} \right] +$$

$$\left.\left. + \beta_i \log(\langle a^i,y\rangle + \sqrt{\langle a^i,y\rangle^2 + 4\alpha_i\beta_i}) \right\} + \text{constant} \right\} .$$

Since, for $\alpha, \beta > 0$,

$$\frac{x + \sqrt{x^2 + 4\alpha\beta}}{2\alpha} > 0 \qquad \forall x ,$$

the dual of (3.8) is unconstrained.

Example 4. The primal problem is

$$(3.9) \quad \begin{cases} \inf \ - \sum_{i=1}^n c_i x_i^{-\beta_i} & c_i > 0,\ 0 < \beta_i < 1 \\ \text{s.t.} \ Ax = b & \\ \quad\ x \geqslant 0 & \end{cases}$$

Denoting

$$q_i(x) = \{-c_i x_i^{-\beta_i} \ , \ x \geqslant 0\},$$

we have

$$q_i^*(x^*) = \begin{cases} -(-x_i^*)^{\frac{\beta_i}{\beta_i - 1}} \cdot (c_i\beta_i)^{\frac{1}{1-\beta_i}} \cdot \left(\frac{1}{\beta_i} - 1\right) & \text{if} \quad x^* < 0 \\ +\infty & \text{if} \quad x^* \geqslant 0 , \end{cases}$$

and the dual of (3.9) is

$$(3.10) \begin{cases} \sup \left\{ \langle y,b \rangle + \sum_{i=1}^{n} (c_i\beta_i)^{\frac{1}{1-\beta_i}} \cdot \left(\frac{1}{\beta_i} - 1\right)(-\langle a^i, y\rangle)^{\frac{\beta_i}{\beta_i - 1}} \right\} \\ \text{s.t} \quad A^T y < 0. \end{cases}$$

Note, that (3.10) is essentially unconstrained, since its feasible set is open, whereas the feasible set of problem (3.5) is closed. This shows that despite the resemblance between problems (3.4) and (3.9), they are of quite a different type.

4. Essentially Unconstrained Duals

A closer examination of the above examples reveals, that the reason for Examples 2 and 3 (Problems (3.6) and (3.8)) to have an unconstrained dual, is that f^* is finite throughout R^n, i.e.

$$(4.0) \qquad \text{dom } f^* = R^n .$$

Here dom f denotes the effective domain of f:

$$\text{dom } f = \{x: f(x) < +\infty\}.$$

This implies that, for the dual objective function h, $\mathrm{dom}\ h = R^m$. Condition (4.0) in turn, is a consequence of the fact that the gradient ∇f maps the interior of $\mathrm{dom}\ f$ *onto* R^n.

This is not the case with Examples 1 and 4 (Problems (3.4) and (3.9)). However, they also belong to a class of problems for which the dual is essentially unconstrained. Moreover, the duality theory corresponding to this class is more complete and closer to the duality theory of linear programming, than the general theory.

One of the common features of the problems in the above mentioned class is that the objective function of (P) is *essentially smooth*.

Definition 4.1: A convex function is essentially smooth, if it satisfies the following three conditions:

(4.1) $\quad \mathrm{int}\,(\mathrm{dom}\ f) \neq \emptyset$;

(4.2) $\quad f$ is differentiable throughout $\mathrm{int}(\mathrm{dom}\ f)$;

(4.3) $\quad \|\nabla f(x_i)\| \to \infty$ for every sequence

$$x_i \to x \in \mathrm{bd}(\mathrm{dom}\ f) \ ,$$

where $\mathrm{bd}\ S$ denotes the boundary of the set S.

In particular, f is essentially smooth if it is smooth (i.e., finite and differentiable throughout R^n), in which case (4.3) holds vacuously.

The following are examples of essentially smooth functions:

a) the objective functions in the four examples of Section 3, i.e.

(i) $\quad \sum_{i=1}^{n} c_i x_i^{-\beta_i} \qquad \beta_i,\ c_i,\ x_i > 0$

(ii) $\sum_{i=1}^{n} x_i \log(x_i/c_i) \qquad c_i > 0, \quad x_i \geq 0$.

(iii) $\sum_{i=1}^{n} \left(\frac{1}{2} \alpha_i x_i^2 - \beta_i \log x_i \right) \qquad \alpha_i, \beta_i, x_i > 0$.

(iv) $- \sum_{i=1}^{n} c_i x^{\beta_i}, \; c_i > 0, \quad 0 < \beta_i < 1, \quad x_i \geq 0$.

b) $\sum_{i=1}^{n} (x_i^{p_i} - \log x_i) \qquad p_i > 1, \quad x_i > 0$.

c) $\sum_{i=1}^{n} x_i \log x_i + h_i(x_i) + a_i x_i \qquad x_i > 0$,

with $h_i(x)$ twice differentiable and satisfying

$$\frac{1}{x} + h_i''(x) > 0$$

$$h_i(0) < \infty \;,$$

$$h_i'(0) < \infty \;,$$

$$h_i'(+\infty) > -\infty \,.$$

Examples for such h_i are

(i) $\alpha_i e^{-\beta_i x_i} \;, \qquad \alpha_i, \beta_i > 0$

(ii) $\frac{1}{2} x_i^2 \log x_i + \alpha_i x_i^2 \qquad$ with $\alpha_i \geq -\frac{5}{4}$

d) $f(x) = \sum_{i=1}^{n} \int_0^{x_i} \log\log(t+1)\,dt, \qquad x_i \geq 0$

e) $f(x) = \sum_{i=1}^{n} \alpha_i x_i^2 + \beta_i x_i \;, \qquad \alpha_i > 0$

f) $f(x) = -\sum_{i=1}^{n} \alpha_i \log x_i, \quad \alpha_i > 0, \; x_i > 0$

g) $f(x) = \sum_{i=1}^{n} \alpha_i x_i \operatorname{tg} x_i, \quad \alpha_i > 0, \; -\frac{\pi}{2} < x_i < \frac{\pi}{2}$

h) $f(x) = \sum_{i=1}^{n} \alpha_i (1 - x_i^{2P_i})^{\frac{1}{2P_i}}, \quad \alpha_i < 0, \; P_i \geq 1$ integer,

$$-1 \leq x_i \leq 1.$$

The following proposition relates essentially smooth functions to essentially unconstrained problems.

Proposition 4.2. Let f be an essentially smooth convex function. Then the problem

$$(4.4) \quad \begin{cases} \inf \quad f(x) \\ x \in \operatorname{dom} f \end{cases}$$

is essentially unconstrained.

Proof. We have to show that every solution of (4.4) is a critical point of f.

Let $\{x_i\}$ be any sequence such that $x_i \in \operatorname{int}(\operatorname{dom} f)$, $x_i \to \bar{x} \in \operatorname{bd}(\operatorname{dom} f)$, and define

$$\alpha = \lim_{i \to \infty} f(x_i) .$$

We establish the existence of a point

$$(4.5) \quad y \in \operatorname{int}(\operatorname{dom} f), \text{ with } f(y) < \alpha .$$

Consider first the one dimensional case. Assume that $\text{int}(\text{dom } f) = (a,b)$, and $\bar{x} = b$. Then b is finite, and $f'(x)$ must tend to $+\infty$ as $x_i \to \bar{x} = b$, by (4.3), since $f'(x)$ is increasing. Let $x^+ \in (a,b)$ be a point with $f'(x) > 0$. Then $f(x)$ is strictly increasing on $[x^+,b)$. Denoting $y = \frac{1}{2} b + \frac{1}{2} x^+$, we have by convexity of f

$$f(y) \leq \frac{1}{2} f(b) + \frac{1}{2} f(x^+)$$

hence

$$f(b) \geq 2f(y) - f(x^+) = f(y) + [f(y) - f(x^+)] > f(y) \tag{4.6}$$

which implies (4.5). Applying the same argument at the left-hand side boundary point a, (this time $f'(x) \to -\infty$ as $x \to a$), completes the proof for the one dimensional case.

In the n-dimensional case, define

$$g(t) = f(x_o + t(\bar{x} - x_o))$$

where x_o is any point in $\text{int}(\text{dom } f)$ (which is not empty by (4.1)). Now g is a one-dimensional essentially smooth function, and the above argument shows that a point y satisfying (4.5) exists, which implies that $\{x_i\}$ is not an infinimizing sequence.

□

Three important properties of essentially smooth functions are stated below (see [9, §26]). We recall that a convex function is called *closed* if it is lower semi-continuous.

Theorem 4.3: Let f be an essentially smooth function which is strictly convex in $\operatorname{int}(\operatorname{dom} f)$. Then

(i) f^* is an essentially smooth closed function, which is strictly convex in $\operatorname{int}(\operatorname{dom} f^*)$,

(ii) ∇f is a one-to-one mapping from $\operatorname{int}(\operatorname{dom} f)$ onto $\operatorname{int}(\operatorname{dom} f^*)$, continuous in both directions ,

(iii) $\nabla f^* = (\nabla f)^{-1}$.

The symmetry in the above theorem may suggest that if the objective function f, of the primal problem (P), is an essentially smooth convex function, so is f^*, and consequently (D) is also an essentially unconstrained problem. This is, however, not enough as shown by the following example, since (P) restricts x additionally from $\operatorname{dom} f$ to $Ax = b$.

Example. Consider the primal problem

$$\inf \frac{1}{4x_1} + \frac{1}{4x_2}$$

$$\text{s.t.} \quad -x_1 + x_2 = 1$$

$$x_1, x_2 \geq 0 .$$

Its dual according to (3.5) is the one-dimensional problem

$$\max \quad y + \sqrt{y} + \sqrt{-y}$$

$$\text{s.t.} \quad -y \leq 0$$

$$y \leq 0 ,$$

which evidently is *not* essentially unconstrained, since its feasible set is the singleton $y = 0$, which is not a critical point of the objective function. We now give a sufficient condition for (D) to be essentially unconstrained.

We denote by $N(A)$ the null-space of the matrix A:

$$N(A) = \{x: Ax = 0\}.$$

When a vector x is orthogonal to a set S, i.e.,

$$\langle x,y\rangle = 0 \quad \forall\, y \in S,$$

we will denote this shortly by $x \perp S$. The set of all vectors orthogonal to S will be denoted $S^{\perp}$.

<u>Theorem 4.4</u>: Assume that in the primal problem (P),

(i) f is essentially smooth and strictly convex in int(dom f).

(ii) $\exists\, x_o \in \text{int(dom } f) \ni \nabla f(x_o) \perp N(A)$.

Then the dual problem (D) is essentially unconstrained.

<u>Proof</u>. By Proposition (4.2), the assertion will be proved if we show that

$$h(y) = \langle y,b\rangle - f^*(A^T y)$$

is essentially smooth.

Since $N(A)^{\perp}$ = range of A^T, it follows from (ii) that $\nabla f(x_o)$ is in the range of A^T, i.e., there exists $y_o \in R^m$, such that

$$\nabla f(x_o) = A^T y_o \quad . \tag{4.7}$$

By Theorem 4.3, there exists x^*, such that

$$\nabla f(x_o) = x^* \in \text{int(dom } f^*). \tag{4.8}$$

Hence there exists an $\varepsilon_o > 0$, such that for each $0 < \varepsilon < \varepsilon_o$ we have

$$A^T(y_o + \varepsilon B) = A^T y_o + \varepsilon A^T B = x^* + \varepsilon A^T B \subset \text{int(dom } f^*) \tag{4.9}$$

where B is the unit ball in R^m:

$$B = \{x \in R^m: \|x\| \leq 1\}.$$

From (4.9) we see that $y_o \in \text{int}(\text{dom } h)$, so that h satisfies (4.1). Since

$$\nabla h(y) = b - A \nabla f^*(A^T y)$$

(where $\nabla f^*(A^T y)$ denotes the gradient of $f^*(\cdot)$, evaluated at $A^T y$), h also satisfies (4.2) and (4.3), i.e., h is essentially smooth.

□

Condition (ii) in Theorem 4.4 will play a central role in the subsequent sections. It will be termed "the orthogonality condition".

A simple case for which this condition holds, *independently of the matrix A* is when f possesses a critical point.

5. Duality Theory

In this section we study the program

(P) $\inf\{f(x): Ax = b, \ x \in \text{dom } f\}$

where f is a closed, essentially smooth function, which is strictly convex on $\text{int}(\text{dom } f)$.

Problem (P) is called *consistent* if it has a feasible solution, and *superconsistent* if

(5.1) $\quad \exists\, \hat{x} \in \text{int}(\text{dom } f) \quad$ such that $A\hat{x} = b$.

The dual of (P), by Section 2, is

(D) $\sup\{\langle b,y\rangle - f^*(A^T y)\}$.

This problem is *superconsistent* if

(5.2) $\exists\, \hat{y}$ such that $A^T\hat{y} \in \operatorname{int}(\operatorname{dom} f^*)$.

Condition (5.2) in turn is equivalent to the previously used *orthogonality condition:*

(5.3) $\exists\, x_o \in \operatorname{int}(\operatorname{dom} f)$ such that $\nabla f(x_o) \perp N(A)$.

On $\operatorname{int}(\operatorname{dom} f^*)$, the conjugate function coincides with its Legendre transform [9, §26] i.e.,

$$L(u) = \langle(\nabla f)^{-1}(u), u\rangle - f((\nabla f)^{-1}(u))$$

and under condition (5.3), $\operatorname{cl} L(A^Ty) = f^*(A^Ty)$, where $\operatorname{cl} L(u) = \liminf_{v \to u} L(v)$ (see [9, Theorem 9.5]). Therefore, under the orthogonality condition, the dual problem of (P) is

$$\text{(D)}\quad \sup\{\langle b,y\rangle - \langle(\nabla f)^{-1}(A^Ty), A^Ty\rangle + f((\nabla f)^{-1}(A^Ty))\}$$

$$\text{s.t.}\quad A^Ty \in \text{range } \nabla f\ .$$

Our first duality result will show that the orthogonality condition (5.3) guarantees the lack of duality gap, and is a sufficient and 'almost' necessary condition for the attainment of the infimum of Problem (P).

<u>Theorem 5.1</u>: If (P) is consistent and satisfies the orthogonality condition, then the infimum of (P) is attained, and is equal to the supremum of (D). Conversely, if the infimum of (P) is attained at an interior point of dom f, then the orthogonality condition holds.

Proof. From (2.6) we have

$$\mathrm{dom}\ g_* = R^m.$$

Also, the existence of a vector $y \in R^m$ such that

$$A^T y \in \mathrm{int}(\mathrm{dom}\ f^*)$$

was demonstrated in Theorem 4.4 (see (4.7) and (4.8)), and the first part of the theorem follows from Fenchel-Rockafellar duality theorem [9, Corollary 31.2.1].

To prove the second part of the theorem, assume that at the minimum point $\hat{x}$, $\nabla f(\hat{x})$ is not orthogonal to $N(A)$. Then there exists a point $d \in N(A)$ such that $\langle d, \nabla f(\hat{x}) \rangle \neq 0$. Replacing d by $-d$ if necessary, we see that d satisfies

$$d^T \nabla f(\hat{x}) < 0 \tag{5.4}$$

$$Ad = 0\ . \tag{5.5}$$

Define $x = \hat{x} + \alpha d$, $\alpha \geqslant 0$. From (5.5) it follows that x is a feasible point for (P). But (5.4) implies (using the Taylor series for f), that

$$f(x) < f(\hat{x})$$

for small enough α, contradicting the optimality of $\hat{x}$. Hence (5.3) must hold at $\hat{x}$, and the theorem is proved. □

The next result, shows that the superconsistency condition (5.1) is not only *sufficient* for the attainment of the supremum in the dual (which is a standard result), but, whenever (5.3) holds it is also a *necessary* condition.

<u>Theorem 5.2:</u> If (P) is superconsistent, then the supremum of the dual problem (D) is attained. Conversely, if the supremum is attained, and (5.3) holds, then (P) is superconsistent.

<u>Proof</u>. Recalling the definition of g:

$$g = -\delta(y|\{b\})$$

we see that $\mathrm{ri}(\mathrm{dom}\ g) = \{b\}$. Hence if (5.1) holds, the supremum of (D) is attained by Fenchel-Rockafellar duality theorem.

Conversely, assume that the supremum of (D) is attained at $\bar{y}$, and (5.3) holds. By Theorem 4.4, (D) is essentially unconstrained, and the gradient of the objective function h of (D) must satisfy at $\bar{y}$

$$\nabla h(\bar{y}) = b - A\,\nabla f^*(A^T\bar{y}) = 0. \tag{5.6}$$

Define

$$x^* = A^T\bar{y}$$

$$\hat{x} = \nabla f^*(x^*)\ . \tag{5.7}$$

Since f^* is essentially smooth, $x^* \in \mathrm{int}(\mathrm{dom}\ f^*)$, hence by (5.5) we have $\hat{x} \in \mathrm{int}(\mathrm{dom}\ f)$, which upon combining (5.6) and (5.7) completes the proof.

□

For dual convex programs, it is well known from the *weak duality* relation

$$\inf\ (P) \geq \sup\ (D)$$

that feasibility of one problem *implies* the boundedness of its dual. The converse is generally false, Linear Programming being a notable

exception. However, for a special class of problems of the type (P) which still includes the four examples of Section 3, we will demonstrate a one-to-one correspondence between primal-feasibility and dual-boundedness.

<u>Theorem 5.3</u>: Consider a primal problem (P), with separable objective function

$$f(x) = \sum_{i=1}^{n} f_i(x_i) \quad ,$$

which satisfies the orthogonality condition.

Then (P) is feasible if and only if (D) is bounded.

<u>Proof</u>. We distinguish two cases (i) $\text{cl}\ (\text{dom}\ f_i) = [a,\infty)$ or $(-\infty,b]$, (ii) $\text{dom}\ f_i = (-\infty,\infty)$. Case (i) can be handled, without loss of generality, by taking $\text{cl}\ (\text{dom}\ f_i) = [0,\infty)$.

If (P) is feasible, then (D) is bounded by the above mentioned weak duality relation.

Let (P) be infeasible, i.e., in case (i) the linear system

(5.8) $\quad Ax = b, \quad x \geqslant 0$

is inconsistent, and in case (ii) the linear system

(5.8') $\quad Ax = b$

is inconsistent.

Then for case (i), by Farkas' Lemma,

(5.9) $\quad \exists\ \bar{y} \ni A^T\bar{y} \leqslant 0, \quad \langle b,\bar{y}\rangle > 0$,

and for case (ii)

(5.9') $\quad \exists\ \bar{y} \ni A^T\bar{y} = 0, \quad \langle b,\bar{y}\rangle > 0$.

By taking $y = M\bar{y}$ (M an arbitrarily large positive scalar), the first term in the dual function

$$h(y) = \langle b,y\rangle - f^*(A^T y)$$

can be made arbitrarily large. For case (ii), the second term is $f^*(0) > -\infty$, which establishes the unboundedness of $h(y)$. For case (i), in order to show that $h(y)$ is unbounded above, it suffices to show, by (5.9) that

$$\lim_{\substack{u\le 0\\ \|u\|\to\infty}} f^*(u) < \infty$$

which reduces to

$$(5.10)\qquad \lim_{\alpha\to-\infty} f_i^*(\alpha) < \infty .$$

Here

$$f_i^*(\alpha) = \alpha f_i'(\alpha)^{-1} - f(f_i'(\alpha)^{-1}) \quad ,$$

thus, setting $t = f_i'(\alpha)^{-1}$, it follows that (5.10) is equivalent to

$$(5.11)\qquad \lim_{t\to 0^+} \{t\, f_i'(t) - f_i(t)\} < \infty \quad .$$

But, since f_i is essentially smooth, with cl dom $f_i = R_+$, it follows that $\lim_{t\to 0^+} f_i'(t) = -\infty$ and $\lim_{t\to 0^+} f_i(t) > -\infty$. Hence (5.11) indeed holds.

6. Examples 1-4. Revisited

In this section we will show that for the Examples introduced in Section 3, the duality theory developed in Sections 4 and 5 is valid to its full extent. This means that

(1) the dual is essentially unconstrained ,

(2) there is no duality gap,

(3) the infimum of the primal is attained,

(4) the supremum of the dual is attained if and only if the primal is superconsistent,

(5) primal feasibility is equivalent to dual boundedness.

For this purpose we have only to show that the orthogonality condition holds for these examples.

Example 1: For problem (3.3) we have

$$f(x) = \sum_{j=1}^{n} f_j(x_j),$$

with

$$f_j(x_j) = L_j\, x_j^{-\beta}\,, \qquad L_j > 0, \;\; \beta > 0$$

and

$$\operatorname{dom} f = \{x \in R^n : x > 0\}.$$

In the pipe networks example (3.1), it was shown that the matrix A is non-negative, with at least one positive entry in each column. In particular then,

$$(6.1) \qquad \sum_{i=1}^{p} a_{ij} > 0 \qquad\qquad j = 1,\dots,n\,.$$

We show that the orthogonality condition holds in this example, i.e.,

$$(6.2) \qquad \exists\, \bar{x} > 0 \;\ni\; \langle \nabla f(\bar{x}), u\rangle = 0 \quad \forall\, u \ni Au = 0.$$

Indeed, take

$$\bar{x}_j = \left(\frac{\beta L_j}{\sum_{i=1}^{p} a_{ij}} \right)^{\frac{1}{1+\beta}} \qquad j = 1,\ldots,n\,,$$

then, $\bar{x} > 0$, and

$$\langle \nabla f(\bar{x}), u\rangle = -\sum_{j=1}^{n} u_j \sum_{i=1}^{p} a_{ij} = \sum_{i=1}^{p} \sum_{j=1}^{n} u_j a_{ij}$$

and the last sum is clearly zero for any u satisfying $Au = 0$.

For the more general problem (3.4), it was assumed that A has a positive row. Let $(a_1,\ldots,a_n)$ be this row, then (6.2) is here satisfied by

$$\bar{x}_j = \left(\frac{a_j}{c_j \beta_j} \right)^{\frac{1}{1+\beta_j}} \quad .$$

Indeed,

$$\langle \nabla f(\bar{x}), u\rangle = -\sum_{j=1}^{n} a_j u_j = 0,$$

since it is one of the equations of the system $Au = 0$.

<u>Example 2:</u> Here

$$f(x) = \sum_{j=1}^{n} f_j(x_j)\,,$$

with

$$f_j(x_j) = x_j \log\left(\frac{x_j}{c_j}\right) \qquad c_j > 0$$

$$\operatorname{dom} f = \{x \in R^n : x \geqslant 0\}\,.$$

Here also, the orthogonality condition is (6.2). It is satisfied trivially by

$$\bar{x}_j = \frac{c_j}{e}$$

(where e is the natural logarithm base) since $\bar{x} > 0$, and

(6.3) $\nabla f(\bar{x}) = 0.$

Example 3: For the objective function here, we have

$$f(x) = \sum_{j=1}^{n} f_j(x_j) ,$$

$$f_j(x_j) = \frac{1}{2}\alpha_j x_j^2 - \beta_j \log x_j \qquad \alpha_j, \beta_j > 0$$

$$\text{dom } f = \{x \in R^n : x > 0\}.$$

As in Example 2, we have for

$$\bar{x}_j = \sqrt{\frac{\beta_j}{\alpha_j}} , \qquad j = 1,\ldots,n ,$$

$\bar{x} > 0$, and

$$\nabla f(\bar{x}) = 0,$$

so (6.2) holds. Note that, in the last two examples, the validity of the orthogonality condition was independent of the matrix A.

Example 4: Here $\text{dom } f^*$ is open:

$$\text{dom } f^* = \{x \in R^n : x < 0\},$$

hence $\text{int}(\text{dom } f^*) = \text{dom } f^*$

and the effective domain of the dual objective function is the open set

$$\text{dom } h = \{y \in R^m, \ A^T y < 0\} \ ,$$

which means in particular that

$$\lim_{u \to y \in \text{bd}(\text{dom } h)} h(u) = -\infty \quad .$$

Therefore $h(y)$ is a barrier function for the set dom h, and the dual problem is essentially unconstrained. Note that by the equivalence between the superconsistency condition (5.2), and the orthogonality condition (5.3), we have that the orthogonality condition is equivalent to dual feasibility.

REFERENCES

[1] Ben-Tal, A., and Charnes, A., "A Dual Optimization Framework for Some Problems of Information Theory and Statistics", Problems in Information and Control, (to appear).

[2] Charnes, A., and Cooper, W.W., "An Extremal Principle for Accounting Balance of a Resource-Value Transfer Economy: Existence, Uniqueness and Computations", Rend. Acad. Naz. Lincei, April 1974, pp. 556-561.

[3] Charnes, A., and Cooper, W.W., "Constrained Kullback-Leibler Estimation; Generalized Cobb-Douglas Balance, and Unconstrained Convex Programming", Rend. Acad. Naz. Lincei, April 1975, pp. 568-576.

[4] Charnes, A., Raike, W.M., and Bettinger, C.O., "An Extremal and Information-Theoretic Characterization of Some Interzonal Transfers", Socio-Econ. Plan. Sci., No.6, 1972, pp. 531-537.

[5] Geoffrion, A.M., "Duality in Nonlinear Programming: a Simplified Application-Oriented Development", SIAM Review, Vol. 13, No.1, 1971.

[6] Khinchin, A.I., Mathematical Foundations of Statistical Mechanics, (translated by G. Gamow), Dover Publications, Inc., New York, 1949.

[7] Kullback, S., and Leibler, R.A., "On Information and Sufficiency", Ann. Math. Stat., Vol. 22 (1951), pp. 79-86.

[8] Murtagh, B.A., "An Approach to the Optimal Design of Networks", Chemical Engineering Science, Vol. 27, 1972, pp. 1131-1141.

[9] Rockafellar, R.T., Convex Analysis, Princeton University Press, Princeton, New Jersey, 1970.

[10] Rothfarb, R., Frank, H., and Rosenbaum, D.M., "Optimal Design of Offshore Natural-Gas Pipeline Systems", Oper. Res., Vol. 18, 1970.

[11] Stoer, J., and Witzgall, C.J., Convexity and Optimization in Finite Dimensions, Springer-Verlag, 1970.

[12] Ben-Israel, A., Charnes, A., and Kortanek, K.O., "Asymptotic Duality Over Closed Convex Sets," J. Math. Anal. Applic., Vol. 35, 1971, pp. 677-691.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) The Center for Cybernetic Studies The University of Texas at Austin	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

(6) A Duality Theory for a Class of Problems with Essentially Unconstrained Duals,

4. DESCRIPTIVE NOTES (Type of report and inclusive dates) (9) Research rept.

5. AUTHOR(S) (First name, middle initial, last name)

(10) Aharon Ben-Tal, Y. Barzilai, Abraham Charnes

6. REPORT DATE (11) January 1979	7a. TOTAL NO. OF PAGES 31	7b. NO. OF REFS 11 (17)
8a. CONTRACT OR GRANT NO. (15) N00014-75-C-0616, & 0569 b. PROJECT NO. NR047-021 N00014-75-C-0569 c. d. (12) 34p.	9a. ORIGINATOR'S REPORT NUMBER(S) CCS- Center for Cybernetic Studies Research Report Number 330 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research (Code 434) Washington, D.C.

13. ABSTRACT

The paper introduces a class of linearly constrained convex programs whose duals are unconstrained in the sense that their solution must be in the interior of the feasible region. A complete duality theory is developed for these problems. Several examples are discussed.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

406 197

Unclassified

Security Classification

A-31408

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Convex Programming						
Duality						